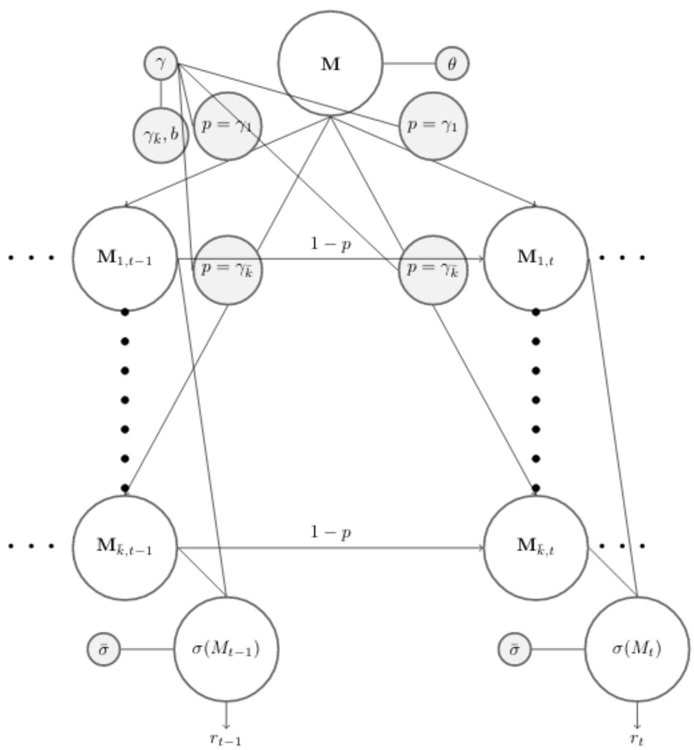
Extending the univariate MSM with Infinite Factorial HMM

Jan Greve, Jaeyeon Lee

August 31, 2018

1 Review of the MSM

Markov Switching Multifractal (in short MSM) is a factorial hidden Markov model with $\bar{k}$ underlying Markov Chains as shown in the graphical model below.



As we can see, each $M_{1,t}, ..., M_{\bar{k},t}$ are independent first order Markov switching processes. Specifically, the transition of each process is governed by the following rule.

$$M_{k,t} = \begin{cases} m \sim M & \text{with probability } \gamma_k \\ M_{k,t-1} & \text{with probability } 1 - \gamma_k \end{cases}$$

where the switching probability γ_k is specified as

$$\gamma_k = 1 - (1 - \gamma_1)^{b^{k-1}}.$$

The state which yields the observation r_t with Gaussian innovation is constructed as follows:

$$\sigma(M_t) \equiv \bar{\sigma}\Big(\prod_{k=1}^{\bar{k}} M_{k,t}\Big)^{\frac{1}{2}},$$

2 Issues with the original construction specific to our framework

The choice of $\bar{k}$ is considered a model selection problem. Generally, the bigger the $\bar{k}$ is, the more variation it can explain. However, since the volatility of the data is bounded in some extent, there should be an appropriate range of $\bar{k}$ and picking a value of $\bar{k}$ above certain threshold will not improve the fit.

In original bivariate construction, only the dependence of switches for same frequency k across series was modeled. Thus, it made sense to pick the same value of $\bar{k}$ to estimate both series and later couple together for the joint estimation.

However, since we want the model to be more flexible and allow to capture conditional dependence of switches in any frequency across series, there is no need to pick the same $\bar{k}$ to fit the model for each series we later jointly model together. Rather, it would be of our interest to infer the most appropriate $\bar{k}$ for each respective series so that our result (inter-frequency dependence structure, forecasting of these series etc.) will not depend on the particular choice of this value, thus greatly reducing the number of experiments we do.

3 Infinite Factorial Hidden Markov Model

Infinite Factorial Hidden Markov Model (IFHMM) developed by Gael et al (2009)[1] allows the number of Markov Switching compoeenents to potentially be unbounded and will infer appropriate range of this in every iteration of the sampler.

The core of this model is a Markovian version of the Indian Buffet Process (IBP) which has an ideal characteristics to be utilized in extending the MSM to allow for not specifying the $\bar{k}$ beforehand.

3.1 Stick breaking construction of the Indian Buffet process

The stick-breaking representation of the IBP developed by Teh et al (2007)[2] serves as a framework for its Markovian version utilized in IFHMM. Here, we will briefly go over characteristics and inference procedure of IBP which are relevant to IFHMM.

IBP is a Bayesian nonparametric distribution over binary matrices with unbounded number of columns. It is typically used as a prior to model objects (corresponds to the row) with unknown number of latent features it possesses (corresponds to the column). It is somewhat similar to a Chinese restaurant process (CRP) but allows customer to belong to multiple tables. The analogous of the IBP to the CRP will prove to be useful in understanding IBP and will occasionally be made here.

Let's start with the definition of the IBP first. We denote the random $N \times K$ binary matrix ($K \to \infty$) generated by the IBP as Z and entry of Z as z_{ik}. Each column which represents the feature has a prior probability μ_k to be active which itself is also given a Beta prior as shown below.

$$\mu_k \sim \text{Beta}(\frac{\alpha}{K}, 1)$$
$$z_{i,k}|\mu_k \sim \text{Bernoulli}(\mu_k)$$

Then, the following observation model with parameters θ, Base measure H and the data x completes the definition.

$$\theta_k \sim H$$
$$x_i \sim F(z_{i,:}, \theta_{:})$$

The stick-breaking construction of the IBP will induce ordering of $\mu_{1:K} = \{\mu_1, ..., \mu_K\}$ as follows:

$$\mu_{(1)} > \mu_{(2)} > \cdots > \mu_{(K)}$$

The distribution of $\mu_{(k)}$ is

$$\nu_{(k)} \sim \text{Beta}(\alpha, 1) \quad \mu_{(k)} = \nu_{(k)}\mu_{(k-1)} = \prod_{l=1}^{k} \nu_{(l)}$$

which compared with the stick-breaking representation of the Dirichelet Process (the parameter for the pmf denoted as $\pi_{(k)}$ and other parameters are the same as above) below

$$\pi_{(k)} = (1 - \nu_{(k)})\mu_{(k-1)} = (1 - \nu_{(k)}) \prod_{l=1}^{k} \nu_{(l)}$$

clarifies that the proportion of the stick discarded from $k-1$ to k represents the pmf of the DP while the one left is that of IBP. The figure below from Teh et al (2007)[2] visualizes this relationship.

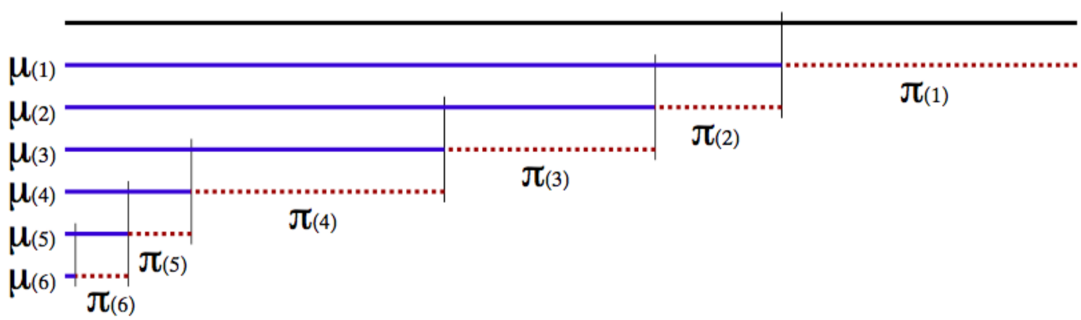


Figure 1: Relationship between weights of the IBP ($= \mu_{(k)}$, blue line) and that of DP ($= \pi_{(k)}$,dotted line)

The exponentially decreasing nature of $\mu_{(k)}$ is critical in modeling state transition probability denoted as γ_k in the original model which we will discuss in later sections.

Additionally, the distribution of $\nu_{(k)}$ can be other distributions such as Pitman-Yor extensions discussed in Teh and Gorur(2009)[3] which induces power-law distributed weights that may align well with the original intention of building MSM which was to represent volatility in scale invariant power-law form.

3.2 Specification of the Markov IBP

The Markov IBP(mIBP) proposed in Gael et al (2009) yields a matrix whose rows represent time steps and columns Markov Switching components of the factorial HMM.

Let's follow the original notation and represent the matrix as S and its indexes as (t,m) where t ranges from time 1 to T while m is potentially unbounded. Each binary Markov switching chain has the following transition matrix

$$W^{(m)} = \begin{pmatrix} 1 - a_m & a_m \\ 1 - b_m & b_m \end{pmatrix}$$

where parameters, initial values of the matrix and its sampling distribution is defined as follows:

$$a_m \sim \text{Beta}(\frac{\alpha}{M}, 1)\ ,\ b_m \sim \text{Beta}(\gamma, \delta)$$
$$s_{0m} = 0\ ,\ s_{tm} \sim \text{Bernoulli}(a_m^{1-s_{t-1,m}} b_m^{s_{t-1,m}})$$

The distributional form of the mIBP prior is computed similary to that of IBP by considering the limit for the left-ordered equivalence class of matricies. For conciseness we don't provide this formula here.

The observation equation of this matrix is very similar to that of the IBP.

$$\theta_k \sim H$$
$$x_i \sim F(s_{t,:}, \theta_:)$$

3.3 Stick breaking representation of the mIBP

Similary to IBP, there is a stick-breaking construction for the mIBP where we have $a_{(1)} > a_{(2)} > \cdots$. The construction is as follows:

$$a_{(1)} \propto \text{Beta}(\alpha, 1)$$
$$p(a_{(m)}|a_{(m-1)} = \alpha a_{(m-1)}^{-\alpha} a_{(m)}^{\alpha-1} \mathbb{1}(0 \leq a_{(m)} \leq a_{(m-1)})$$

while variables b_m will be drawn independently from $\text{Beta}(\gamma, \delta)$ and reordered in accordance with the value of $a_{(m)}$.

Because of this construction, the resulting matrix S will have exponentially decreasing number of rows activated (at least in expectation) when looking column by column from left to right which aligns with the construction of the original MSM (remember the formula for the γ_k). So far, the only prior for a matrix with varying column size that could induce this sort of sparsity was IBP. Thus, we chose IFHMM which is based on the time series extension of IBP as our choice of model to extend MSM.

3.4 Inference using Slice sampling

Similar to IBP (although we did not cover this part above) we can adaptively truncate the matrix S to do inference with Slice Sampling.

We introduce an auxiliary slice variable μ with the following form

$$\mu \sim \text{Uniform}(0, \min_{m:\exists t, s_{tm}=1} a_m)$$

Then, using this sampled variable, the algorithm proceeds as follows:

1. Sample the slice variable μ with the formula above.
2. Pick columns of S that meets $a_m > \mu$. Extend the column size of S if necessary.
3. With the new S, run forward and backward sampling for each column (columns each independently follows a Markov Chain) to update elements of the S.
4. Resample other model parameters such as $(\theta_:, \alpha, \gamma, \delta)$.

3.5 Observation equation (*still working on it)

In order to conduct these update and sampling from the posterior, we need a concrete form of observation equation to compute the likelihood.

The truncated form of the mIBP we obtain in each iteration of the sampler is a matrix with each column representing a first order binary Markov Switching processes.

As mentioned in the proposal, we only consider Binomial MSM which means that the distribution of new draws for $M_{k,t}$ yields a value m_0 or $2 - m_0$ in equal probability (due to the constraint of the MSM to have mean 1 distribution). Thus the transition matrix of frequency k (say $P(\gamma_k)$) of the original model is simply

$$P(\gamma_k) = \begin{pmatrix} 1 - \gamma_k/2 & \gamma_k/2 \\ \gamma_k/2 & 1 - \gamma_k/2 \end{pmatrix}$$

where one row and column represents the state m_0 and another as $2 - m_0$. Thus, we can convert the binary matrix S to this form and sample $\bar{\sigma}$ to construct the state $\sigma(M_t)$ which yields the observed returns r_t with Gaussian innovation. From this, we can compute the likelihood.

Another more general approach is to treat S as matrix of arrivals and non-arrivals. If $s_{tk} = 1$ we treat that $m \sim M$ which is a volatility arrival at time t and if not $M_{k,t} = M_{k,t-1}$ as denoted in the first section. This is a more general setting which can be extended on non-binomial MSM case as well.

4 Future work

The sampling scheme is almost complete and all we need to do now is start to implementing the sampler. We expect there are plenty of unforeseen implementational difficulties ahead for this work alone. Then, we would move on to the main concern of our proposal which deals with the modeling the conditional dependence of these switches.

References

[1] Jurgen V. Gael, Yee W. Teh, and Zoubin Ghahramani. The infinite factorial hidden markov model. *Advances in Neural Information Processing Systems*, pages 1697–1704, 2009.

[2] Yee Whye Teh, Dilan Grür, and Zoubin Ghahramani. Stick-breaking construction for the indian buffet process. *In Artificial Intelligence and Statistics*, pages 556–563, 2007.

[3] Yee W. Teh and Dilan Gorur. Indian buffet processes with power-law behavior. *In Advances in neural information processing systems*, pages 1838–1846, 2009.